

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT: HYEONG-SOP SHIM, ET AL.)
FOR: ROBUST GENOTYPING METHOD USING)
DNA CHIP AND DNA CHIP USED THEREIN)

CLAIM FOR PRIORITY

Mail Stop Patent Application
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Commissioner:

Enclosed herewith is a certified copy of Korean Patent Application No. 2003-0005025 filed on January 25, 2003. The enclosed Application is directed to the invention disclosed and claimed in the above-identified application.

Applicants hereby claim the benefit of the filing date of January 25, 2003, of the Korean Patent Application No. 2003-0005025, under provisions of 35 U.S.C. 119 and the International Convention for the protection of Industrial Property.

Respectfully submitted,

CANTOR COLBURN LLP

By: _____
Soonja Bae

Reg. No. (See Attached)
Cantor Colburn LLP
55 Griffin Road South
Bloomfield, CT 06002
PTO Customer No. 23413
Telephone: (860) 286-2929
Fax: (860) 286-0115

Date: January 23, 2004



별첨 사본은 아래 출원의 원본과 동일함을 증명함.

This is to certify that the following application annexed hereto
is a true copy from the records of the Korean Intellectual
Property Office.

출원번호 : 10-2003-0005025
Application Number

출원년월일 : 2003년 01월 25일
Date of Application JAN 25, 2003

출원인 : 삼성전자주식회사
Applicant(s) SAMSUNG ELECTRONICS CO., LTD.



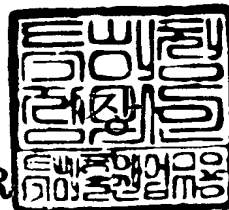
2003 년 02 월 07 일

특

허

청

COMMISSIONER





【서지사항】

【서류명】	특허출원서
【권리구분】	특허
【수신처】	특허청장
【참조번호】	0002
【제출일자】	2003.01.25
【국제특허분류】	C12Q
【발명의 명칭】	D N A 칩을 이용하여 유전자형을 판별하는 강건한 방법 및 이에 사용되는 D N A 칩
【발명의 영문명칭】	Robust genotyping using DNA chip and DNA chip used therein
【출원인】	
【명칭】	삼성전자 주식회사
【출원인코드】	1-1998-104271-3
【대리인】	
【성명】	이영필
【대리인코드】	9-1998-000334-6
【포괄위임등록번호】	2003-003435-0
【대리인】	
【성명】	이해영
【대리인코드】	9-1999-000227-4
【포괄위임등록번호】	2003-003436-7
【발명자】	
【성명의 국문표기】	심형섭
【성명의 영문표기】	SHIM, Hyeong Sop
【주민등록번호】	700726-1156311
【우편번호】	441-390
【주소】	경기도 수원시 권선구 권선동 유원아파트 603-1205
【국적】	KR
【발명자】	
【성명의 국문표기】	박태성
【성명의 영문표기】	PARK, Tae Sung
【주민등록번호】	620319-1840814

【우편번호】	137-030
【주소】	서울특별시 서초구 잠원동 60-3 한신아파트 309-203
【국적】	KR
【발명자】	
【성명의 국문표기】	양재식
【성명의 영문표기】	YANG, Jae Shick
【주민등록번호】	710120-1047422
【우편번호】	305-390
【주소】	대전광역시 유성구 전민동 세종아파트 109-203
【국적】	KR
【심사청구】	청구
【취지】	특허법 제42조의 규정에 의한 출원, 특허법 제60조의 규정에 의한 출원심사를 청구합니다. 대리인 이영필 (인) 대리인 이해영 (인)
【수수료】	
【기본출원료】	20 면 29,000 원
【가산출원료】	12 면 12,000 원
【우선권주장료】	0 건 0 원
【심사청구료】	16 항 621,000 원
【합계】	662,000 원
【첨부서류】	1. 요약서·명세서(도면)_1통

【요약서】**【요약】**

본 발명은 DNA 칩을 이용하여 유전자형을 판별하는 강건한 방법에 관한 것이다. 본 발명에 따른 유전자형 판별 방법은 각 돌연변이 위치 별로 정상형 유전자에 완전히 일치하는 프로브와 돌연변이형 유전자에 완전히 일치하는 프로브로 이루어진 최적 프로브 쌍만이 배열되어 있는 DNA 칩을 사용한다. 먼저, 유전자형을 이미 알고 있는 표준 핵산을 DNA 칩에 혼성화시켜 얻어진 데이터로부터 유전자형 판별 함수를 설정한다. 이어서, DNA 칩에 유전자형을 모르는 타겟 핵산을 혼성화시켜 얻어진 데이터로부터 입력 벡터를 계산하고 이를 판별 함수에 입력하여 타겟 핵산의 유전자형을 판별한다.

본 발명에 따르면, 각 돌연변이 위치 당 최적의 프로브 쌍을 이용하여 타겟 샘플의 유전자형을 강건하게 통계적으로 판별할 수 있다.

【대표도】

도 1

【색인어】

DNA 칩, 유전자형 판별(genotyping), 판별 함수, 로지스틱 회귀분석, 사후 확률

【명세서】**【발명의 명칭】**

DNA 칩을 이용하여 유전자형을 판별하는 강건한 방법 및 이에 사용되는 DNA 칩
{Robust genotyping using DNA chip and DNA chip used therein}

【도면의 간단한 설명】

도 1은 본 발명의 유전자형 판별 방법의 흐름도이다.

도 2는 최적 프로브 쌍 선발 단계의 세부 흐름도이다.

도 3은 유전자형 판별 함수 설정 단계의 세부 흐름도이다.

도 4는 MODY2 alpha 1 유전자의 엑손 2 의 1번 돌연변이 위치(E02-01)에 대한 유전자형 판별 함수 설정을 위해 도시한 MA 그래프이다.

도 5는 유전자형 판별 단계의 세부 흐름도이다.

도 6은 MODY2 alpha 1 유전자의 엑손 4의 18번 돌연변이 위치(E04-18)에 대한 유전자형 판별 함수 설정을 위해 도시한 MA 그래프 위에 유전자형을 모르는 타겟 핵산의 혼성화 결과를 함께 도시한 그래프이다.

도 7 및 도 8은 교차 혼성화 결과를 도시하는 MA 그래프들이다.

【발명의 상세한 설명】**【발명의 목적】****【발명이 속하는 기술분야 및 그 분야의 종래기술】**

<8> 본 발명은 DNA 칩을 사용하여 유전자형을 모르는 타겟 핵산이 정상형인지 돌연변이형인지를 강건하게 판별하는 방법에 관한 것이다.

- <9> 유전자형을 판별 분석하기 위한 가장 보편적인 방법으로는 서열 분석기(sequencing machine)를 사용하여 서열을 확인하는 방법 등이 있다. 그러나, 이 방법은 정확하기는 하나 한번에 여러 샘플을 처리하지 못하기 때문에 수율이 낮다는 단점이 있다.
- <10> 최근에는 동시에 여러 위치의 유전자형을 한꺼번에 판별할 수 있는 DNA 칩이 미국 특허 제 6,027,880호 및 제6,300,063호에 개시되어 많은 사람들의 관심을 끌고 있다. 이들 특허에 개시되어 있는 DNA 칩은 기본적으로는 돌연변이가 있다고 알려진 위치에 A,C,G,T를 가지고 돌연변이가 있지 않은 다른 부위는 모두 같은 염기배열을 가지는 길이 9~25mer의 올리고뉴클레오타이드(oligonucleotide)를 프로브로 사용하는 타일드 어레이(tiled array) 방식을 채용하고 있다. 즉, 서열결정(sequencing)의 기능을 함께 달성하기 위해서 돌연변이가 있다고 알려진 염기 서열의 위치뿐만 아니라 돌연변이 위치의 근처에서도 타일드 어레이 방식으로 가능한 모든 조합의 염기서열을 프로브로 사용하기 때문에 타일드 어레이를 적용하는 위치의 수가 증가함에 따라 필요한 프로브의 숫자도 4배씩 증가한다.
- <11> 그런데, 이미 각 유전자형의 서열이 알려져 있는 경우라면 타일드 어레이 방식은 불필요한 프로브를 칩 위에 올려놓아야 하는 단점이 있다. 또, 타일드 어레이 방식은 삽입(insertion) 이나 결실(deletion)과 같은 돌연변이에는 적용하기가 어렵다. 그리고, 타일드 어레이 방식은 프로브의 길이를 고정시킨 상태에서 유전자 위치(genotype locus)에 일치(match)가 되는 부분을 이동시키면서 매우 유사한 서열을 가진 프로브들을 반복한다. 따라서, 특정 유전자 위치에 대한 유전자형을 판

별 분석하는데 있어서 대량의 프로브들을 사용하기 때문에 결과를 해석하기가 어렵고 DNA 칩의 제조 비용이 상승한다. 예를 들면, 정상형(wild type) 유전자에 완전 일치하는 프로브(perfect match probe)나 돌연변이형(mutant type) 유전자에 완전 일치하는 프로브와의 혼성화 강도(hybridization intensity)보다 나머지 두 개의 불일치 프로브(mismatch probe)와의 혼성화 강도가 높게 나올 경우 유전자형 판별에 오류가 생기고 이로 인해 교차-혼성화(cross-hybridization) 효과를 명확히 구분할 수 없다. 또, 프로브의 길이가 고정되어 있기 때문에 실제 혼성화에 가장 적합한 프로브를 사용할 수 없다는 단점이 있다.

【발명이 이루고자 하는 기술적 과제】

<12> 본 발명이 이루고자 하는 기술적 과제는 프로브의 길이에 제한을 받지 않고 프로브의 수를 최소화하면서도 유전자형 판별시의 오류에 대해서 강건하게 유전자형을 판별할 수 있는 방법을 제공하기 위한 것이다.

<13> 본 발명이 이루고자 하는 기술적 과제는 상기 유전자형 판별 방법에 사용되는 DNA 칩을 제공하기 위한 것이다.

【발명의 구성 및 작용】

<14> 상기 기술적 과제를 달성하기 위한 본 발명에 따른 유전자형의 판별 방법에 따르면, 각 돌연변이 위치 별로 정상형 유전자에 완전히 일치하는 프로브와 돌연변이형 유전자에 완전히 일치하는 프로브로 이루어진 최적 프로브 쌍만이 배열되어 있는 DNA 칩을 사용한다. 바람직하기로는 상기 각 돌연변이 위치 별로 상기 최적 프로브 쌍이 2회 이상 배열된다.



- <15> 먼저, 유전자형을 이미 알고 있는 표준 핵산을 DNA 칩에 혼성화시켜 얻어진 데이터로부터 유전자형 판별 함수를 설정한다. 이어서, DNA 칩에 유전자형을 모르는 타겟 핵산을 혼성화시켜 얻어진 데이터로부터 입력 벡터를 계산하고 이를 판별 함수에 입력하여 타겟 핵산의 유전자형을 판별한다. 유전자형 판별 후에 교차 혼성화 결과를 체크하여 유전자형 판별 결과를 보정한다.
- <16> 기타 실시예들의 구체적 사항들은 상세한 설명 및 도면들에 포함되어 있다.
- <17> 이하 첨부한 도면을 참조하여 본 발명에 따른 유전자형 판별 분석 방법 및 이에 사용되는 DNA 칩에 관한 실시예들을 설명한다. 그러나 본 발명은 이하에서 개시되는 실시예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 것이며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하며, 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다. 명세서 전체에 걸쳐 동일 참조 부호는 동일 부재를 지칭한다.
- <18> 본 명세서에서, DNA 칩은 다량의 핵산 프로브를 마이크로어레이(microarray) 상태로 배열한 것을 지칭한다. 핵산(nucleic acid)은 시토신(C), 티민(T), 우라실(U), 아데닌(A) 및 구아닌(G) 등의 피리미딘 및 퓨린을 포함하는 뉴클레오사이드 또는 뉴클레오타이드의 폴리머 또는 올리고머(폴리뉴클레오타이드 또는 올리고뉴클레오타이드)를 지칭한다. 그 종류로는 최소 500bp 이상의 프로브를 사용하는 cDNA 칩 및 9~25mer의 올리고뉴클레오타이드(oligonucleotide)를 프로브로 사용하는 올리고뉴클레오타이드 칩을 들 수 있다.
- <19> 본 명세서에서 사용하는 "표준 핵산(standard nucleic acid)"은 이미 유전자형이 알려져 있는 핵산을 지칭한다. 그리고 "타겟 핵산(target nucleic acid)"은 유전자형을

알고자 하는 핵산(RNA 또는 DNA의 올리고뉴클레오타이드 또는 폴리뉴클레오타이드)을 지칭한다. 그리고 "프로브"란 타겟 핵산의 유전자형을 판별 분석하기 위하여 사용하는 핵산을 지칭한다. 이하, 정상형 유전자에 완전 일치하는 프로브는 "wp"로 돌연변이형 유전자에 완전 일치하는 프로브는 "mp"로 지칭한다.

<20> 각 흐름도에서 점선으로 표시한 부분은 선택적인(optional) 단계를 나타낸다.

<21> 도 1을 참조하면, 본 발명에 따른 강건한 유전자형 판별 방법은 유전자형 판별 함수 설정(genotyping algorithm set up) 단계(200) 및 타겟 핵산의 유전자형 판별(genotyping) 단계(300)로 구성된다. 필요에 따라서(optionally), 유전자형 판별 함수 설정 단계(200) 전에 최적 프로브 쌍 선발(screening) 단계(100)를 유전자형 판별 단계(300) 후에 유전자형 판별 결과 보정 단계(400)를 더 포함할 수도 있다. 본 발명에 따른 유전자형 판별 방법은 각 돌연변이 위치 별로 정상형 유전자에 완전히 일치하는 프로브와 돌연변이형 유전자에 완전히 일치하는 프로브로 이루어진 최적 프로브 쌍만이 배열되어 있는 DNA 칩을 사용하여 유전자형을 판별한다. 따라서, 불필요한 프로브를 칩 위에 올려놓지 않아도 되고 결과 해석이 간단하며 교차-혼성화에 의한 오류도 쉽게 교정할 수 있으며 DNA 칩의 제조 단가를 낮출 수 있다. 또, 삽입(insertion)이나 결실(deletion)과 같은 돌연변이에도 적용할 수 있다. 이하 각 단계별로 구체적으로 설명한다.

<22> 최적 프로브 쌍 선발

<23> 도 2는 도 1에 도시되어 있는 최적 프로브 쌍 선발 단계(100)의 세부 흐름도이다. 구체적으로, 인 실리코(In silico) 방법에 의한 프로브 설계에 의해 돌연변이가 있다고 알려진 위치 당 복수개의 "wp"와 복수개의 "mp"를 디자인한다(101). "wp"와 "mp"는 동일

길이일 수도 있고 다른 길이일 수도 있다. 즉 "wp"와 "mp"의 길이에겐 제한이 없으며 단지 동일 가닥(strand)이라는 제한만 존재한다. 이어서, 디자인된 복수개의 "wp"와 "mp"를 모두 조합하여 복수개의 "wp-mp 쌍"을 기판에 고정화시켜서 최적 프로브 쌍 선발용 칩을 완성한다(103). "wp-mp 쌍"을 기판에 올리는 방법으로는 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에게 공지된 다양한 방법 중의 하나가 사용될 수 있다. 또, 본 출원인에 의해 출원된 대한민국 특허출원 2001-53687호에 개시되어 있는 방법에 따라서 "wp-mp 쌍"을 칩 위에 올릴 수 있다. 상기 출원의 전체 내용은 참조용으로 본 명세서에 속하는 것으로 한다.

<24> 계속해서, 완성된 칩에 대하여 표준 핵산을 혼성화시킨다(105). 이 때, 복수의 최적 프로브 쌍 선발용 칩에 대하여 혼성화를 진행한다. 혼성화 또한 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 공지된 다양한 방법을 사용하여 진행한다. 혼성화가 완료되면 혼성화 강도 정량 데이터(hybridization intensity quantification data)를 수집한다(107). 혼성화 강도 정량 데이터는 스캐너를 사용하여 수집한다. 복수의 최적 프로브 쌍 선발용 칩으로부터 다수의 혼성화 강도 정량 데이터를 수집한다. 마지막으로, 얻어진 혼성화 강도 정량 데이터로부터 각 돌연변이 위치 당 최적의 "wp-mp 쌍"을 선발(screening)한다(109). 최적의 프로브 쌍 선발을 위한 기준으로는 하기 수학적 식 1과 같은 판별식을 사용하는 것이 바람직하다.

<25> [수학적 식 1]

<26> $\{\text{Mean}(\ln(r^{wt})) - 2 \text{SD}(\ln(r^{wt}))/\sqrt{N^{wt}}\} - \{\text{Mean}(\ln(r^{mt})) + 2\text{SD}(\ln(r^{mt}))/\sqrt{N^{mt}}\}$

- <27> 상기 식에서, N 은 각 표준 핵산과 혼성화를 시행한 횟수이고, r^{wt} 은 정상형 표준 핵산의 w_p 에 대한 혼성화 강도를 정상형 표준 핵산의 m_p 에 대한 혼성화 강도로 나눈 값이고, r^{mt} 는 돌연변이형 표준 핵산의 w_p 에 대한 혼성화 강도를 돌연변이형 표준 핵산의 m_p 에 대한 혼성화 강도로 나눈 값이다. DNA 칩과 표준 핵산과의 혼성화에서 하나의 $\ln(r)$ 값이 얻어지고 혼성화를 N 회 반복하면 $\ln(r)$ 의 분포가 발생한다. Mean은 $\ln(r)$ 값들의 평균값을 SD는 표준편차를 나타낸다. 평균값(Mean) 대신 $\ln(r)$ 값들의 중위수값(Median)을 사용할 수도 있다.
- <28> 상기 수학식 1에 대입하여 계산한 값이 가장 큰 프로브 쌍이 최적의 프로브 쌍으로 선택된다.
- <29> 최적의 프로브 쌍은 본 출원인과 동일 출원인에 의해 2002년 3월 6일자로 출원된 대한민국특허출원번호 02-11871호에 개시되어 있는 방법에 의해 선택될 수도 있다. 상기 출원의 전체 내용은 참조용으로 본 명세서에 속하는 것으로 한다.
- <30> 돌연변이 위치 당 각각 최적의 프로브 쌍이 알려져 있는 경우에는 이 단계를 생략할 수 있음은 물론이다.
- <31> 유전자형 판별 함수 설정
- <32> 도 2의 흐름도에 따라 돌연변이 위치 별로 각각 최적의 프로브 쌍이 선택되면 유전자형 판별 함수를 설정한다.
- <33> 유전자형 판별 함수 설정 단계의 세부 흐름도인 도 3을 참조하면, 먼저 각 돌연변이 위치 별로 최적 프로브 쌍만을 마이크로어레이 상태로 배열하여 DNA 칩을 제조한다(201). 칩의 제조는 최적 프로브 쌍 선발용 칩의 제조와 동일한 방법으로 제조한다. 바

람직하기로는 각 돌연변이 위치 별로 최적의 프로브 쌍을 2회 이상 반복해서 배열하는 것이 유전자형 판별의 품질 관리(Quality Control:QC) 및 품질 보증(Quality Assurance:QA) 측면에서 바람직하다. 가장 바람직하기로는 각 돌연변이 위치 당 wp를 2회 이상 배열한 후, 바로 그 옆에 mp를 동수로 배열하는 것이 육안으로도 혼성화 결과를 식별할 수 있어서 바람직하다. QC 및 QA와 칩 제조 비용의 관점에서 각 돌연변이 위치 당 wp와 mp를 각각 3회 칩 상에 배열하는 것이 가장 바람직하다.

<34> 이어서, 타겟 샘플과 칩을 혼성화시킨(203) 후, 혼성화 강도 정량 데이터를 수집한다.(205). 타겟 샘플과 칩을 혼성화시킨 후, 세정(washing)하여 스캐너로 혼성화 강도를 정량화한다.

<35> 얻어진 혼성화 데이터 중 불량 스팟(bad spot) 데이터를 제거하는 단계를 더 진행할 수도 있다(207). 불량 스팟을 판별하는 기준으로는 다수의 통계적인 데이터 등에 근거하여 확립한 각 스팟 지름(diameter) 유효값의 임계치(cutoff value) 또는 스팟 밝기(intensity) 유효값의 임계치 등을 들 수 있다. 본 발명에서는 각 스팟 지름의 유효값보다 큰 지름을 가지는 스팟을 불량 스팟으로 간주하여 통계 데이터 처리시 제거시켰다.

<36> 계속해서, 얻어진 혼성화 강도 정량 데이터로부터 판별 함수 설정용 벡터(vector)를 계산한다(209). 벡터는 비모수(nonparametric) 통계학에서 전형적으로 사용하는 호지-레만 추정(Hodge-lehman estimation: 이하 H-L 추정) 방법을 적용하여 계산하는 것이 최종적으로 얻어지는 유전자형 판별 함수의 강건도(robustness)를 높이는데 적합하다. 본 발명에 따른 유전자형 판별 함수를 설정하기 위해 사용되는 벡터(vector)는 비율(ration) 성분 및/또는 강도(intensity) 성분으로 구성된다.

<37> 비율 성분은 다음과 같이 구한다.

- <38> 먼저, 하기 수학적 식 2와 같이, 정상형 프로브(w_{pi})의 혼성화 강도와 돌연변이형 프로브(mp_j)의 혼성화 강도의 가능한 모든 조합의 수($i \times j = n$) 만큼의 비율을 계산한다.
- <39> [수학적 식 2]
- <40> $r_{ij} = (w_{pi} \text{의 혼성화 강도} / mp_j \text{의 혼성화 강도})$
- <41> 모든 r_{ij} 값을 계산한 후에 그 값들을 $r(1) \leq r(2) \leq \dots \leq r(n-1) \leq r(n)$ 과 같이 크기 순서대로 배열한 후, 이 중에서 중위수(median)인 $r(m)$ 을 구한다.
- <42> 예컨대, w_p 를 3회 mp 를 3회 배열하였을 때에는 9개의 비율 조합이 나오고 이를 오름차순으로 정렬하면 $r(1) \leq \dots \leq r(5) \leq \dots \leq r(9)$ 가 되고, $r(m)$ 은 $r(5)$ 가 된다.
- <43> 이렇게 구해진 중위수 $r(m)$ 에 하기 수학적 식 3과 같이 자연로그(\ln)를 취한 값을 비율 성분(M)으로 사용한다.
- <44> [수학적 식 3]
- <45> $M = \text{비율 성분} = \ln(r(m))$
- <46> 경우에 따라서는 자연로그(\ln) 대신에 상용로그(\log)를 취한 값을 비율 성분으로 사용할 수도 있다.
- <47> 이와 같이 중위수를 사용하면 동일 프로브에 대한 산술 평균을 구해서 비율 성분을 계산하는 것보다 실험의 오차에 대해 강건해진다. 예컨대, 아래 표 1과 같이 mp_3 의 혼성화 강도 값이 mp_1 및 mp_2 에 비해 차이가 많이 날 경우에는 산술 평균을 사용해서 계산하는 것보다는 H-L 추정에 의해 중위수를 구해서 비율 성분을 계산하는 것이 실험의 오차에 대해서 강건해진다.
- <48> [표 1]

<49> 프로브 이름	혼성화 강도
E04-22-wp1	948
E04-22-wp2	870
E04-22-wp3	1071
E04-22-wp1	74
E04-22-wp2	73
E04-22-wp3	1363

<50> 강도(intensity) 성분은 다음과 같이 구한다.

<51> 첫 번째 방법은 하기 수학식 4와 같이, 정상형 프로브(wp_i)의 혼성화 강도와 돌연변이형 프로브(mp_j)의 혼성화 강도의 가능한 모든 조합의 수($i \times j = n$) 만큼 강도의 곱(a_{ij})을 계산한다.

<52> [수학식 4]

<53> $A = a_{ij} = wp_i \times mp_j$

<54> 모든 a_{ij} 값을 계산한 후에 그 값들을 앞에서 설명한 바와 동일한 방식으로 H-L 추정에 의해 중위수 $a(m)$ 을 구한다.

<55> 이어서, 하기 수학식 5와 같이 중위수 $a(m)$ 에 자연로그(ln)를 취한 후 이를 2로 나눈 값을 강도 성분(A)으로 사용한다.

<56> [수학식 5]

<57> $A = \text{강도 성분} = \ln(a(m))/2$

<58> 두 번째 방법은 하기 수학식 6과 같이 wp_i 와 mp_j 의 모든 조합에 대해 최대수를 구한다.

<59> [수학식 6]

<60> $m_{ij} = \max (wp_i , mp_j)$

- <61> 이렇게 얻어진 m_{ij} 의 중위수 $m(m)$ 를 구하고 하기 수학식 7과 같이 상용로그(log)를 취하여 강도 성분(A)을 구할 수도 있다.
- <62> [수학식 7]
- <63> $A = \text{강도 성분} = \log(m(m))$
- <64> 물론 필요에 따라서, 수학식 5와 7에서 사용하는 로그는 자연로그와 상용로그를 서로 바꾸어 사용할 수도 있다.
- <65> 203 내지 209의 과정을 복수개의 칩에 대하여 진행하여 복수개의 비율 성분(M)들과 강도 성분(A)들을 구한다.
- <66> 상술한 방법들에 의해 혼성화 강도 데이터로부터 비율 성분(M)들과 강도 성분(A)들로 구성된 벡터들을 구한 후에 이들 벡터들을 사용하여 판별 함수를 설정한다(211).
- <67> 판별 함수를 설정하기 위해서는 얻어진 비율 성분(M)들을 y축으로 하고 강도 성분(A)들을 x축으로 하여 MA 그래프(MA plot)를 그린다.
- <68> 도 4는 MODY2 alpha 1 유전자의 엑손(exon) 2 의 1번 위치(E02-01)에 대한 유전자형 판별 함수 설정을 위해 도시한 MA 그래프이다. 도 4의 MA 그래프는 다음과 같은 과정을 거쳐서 얻었다.
- <69> 먼저, E02-01 wp를 3회 기판 위에 올리고 그 옆에 E02-01 mp를 3회 기판 위에 올려서 칩을 완성하였다. 칩은 아민기를 가진 E02-01 wp와 E02-01 mp를 에폭시기를 갖는 폴리에틸렌글리콜(polyethyleneglycol, PEG) 유도체를 이용한 하이드로겔과 혼합한 스포팅 용액을 사용하여 제조하였다. 아민기를 갖도록 표면 처리된 유리 표면에 위와 같이 구성된 스포팅 용액을 바이오로봇 프린터(모델 PixSys 5500, Cartesian Technologies, Inc.,

CA, 미국)를 이용하여 스폿팅한 후, 37℃에서 4시간 동안 습한 인큐베이터에서 방치하였다. 이어 배경 노이즈의 제어에 필요한 공정, 즉, 표준 핵산이 유리 표면에 부착하지 않도록 하기 위해 스폿팅되지 않은 위치의 유리 표면 아민기가 음전하를 띄도록 반응을 실행하고 건조기에 보관하였다.

<70> 표준 핵산에는 형광 물질을 표지하였다. 형광물질로는 FITC, 플루오레세인, Cy3, Cy5 또는 텍사스레드 등이 사용될 수 있으나 도 4의 실험의 경우에는 Cy3-dUTP를 사용하였다.

<71> 표준 핵산과 프로브의 혼성화 조건은 0.1% 6SSPET(0.1% Triton X-100을 포함하는 Saline Sodium Phosphate EDTA Buffer) 용매에 용해된 20nM 농도의 표준 핵산과 칩을 37℃에서 16시간 동안 반응시킨 후, 칩을 0.05% 6SSPET와 0.05% 3SSPET로 각 5분씩 상온에서 세척하고, 5분간 상온에서 건조시킨 후, 스캐닝하였다. 이 때 스캐닝은 액손(Axon) 스캐너(모델 GenePix 4000B, Axon Instrument, Inc., CA, 미국)를 이용하였고, 스캐닝 데이터를 해석하는 프로그램으로는 동사의 진픽스 프로 3.0(GenePix Pro 3.0) 프로그램을 사용하여 비율 성분과 강도 성분을 계산한 후, 도 4의 MA 그래프를 도시하였다. 각 데이터는 정상형 표준 핵산과는 100여장 칩과 혼성화를 통해서, 돌연 변이 표준 핵산과는 30여장 칩과의 혼성화를 통해 얻은 것이다.

<72> 도 4와 같이 강도 성분이 변하더라도 비율 성분이 유지되는 경우에는 비율 성분만을 사용하여 판별함수를 설정한다. 판별함수로는 로지스틱 회귀분석(logistic regression)을 사용하여 추정된 회귀 계수(logistic regression coefficient)(a, b)를 구하여 판별함수를 설정하는 것이 바람직하다.

<73> 한편, 도 4와 달리 강도 성분이 변화할 때(예컨대, 낮아질 때) 비율 성분이 현저히 변화하는 경향(예컨대, 감소하는 경향)을 보일 경우에는, ① 비율 성분이 감소하는 부분 집합을 제외하고 나머지 비율 성분들의 집합을 사용하고 로지스틱 회귀분석(logistic regression)을 판별함수로 사용하여 회귀 계수를 구하거나, ② 비율 성분과 강도 성분을 2차원 벡터로 하여 판별함수를 구할 수 있다. 그러나, 칩의 특성상 강도의 변이 (variation)가 큰 점을 고려하면 ② 번 방법보다는 ① 번 방법이 더 바람직하다.

<74> 유전자형 판별

<75> 상술한 단계들에 의해 유전자형 판별 함수가 설정되면 유전자형을 알고자 하는 타겟 핵산에 대한 유전자형 판별을 실시한다. 유전자형 판별은 타겟 핵산을 DNA 칩에 적용하여 얻어진 테스트 결과로부터 입력 벡터를 계산하여 이를 유전자형 판별 함수 설정 단계로부터 얻어진 판별 함수에 적용함으로써 진행한다.

<76> 유전자형 판별 단계의 세부 흐름도인 도 5에 도시되어 있는 바와 같이, 입력 벡터 계산 단계까지는 유전자형 판별 함수 설정 방법과 동일하게 진행된다. 먼저, 유전자형을 알고자 하는 타겟 핵산과 유전자형 판별 함수가 설정되어 있는 칩을 혼성화를 시킨다 (303). 이어서, 혼성화 강도 정량 데이터를 수집한다(305). 이어서, 얻어진 혼성화 데이터 중 불량 스팟(bad spot) 데이터를 제거하는 단계를 선택적으로 더 진행할 수도 있다 (307).

<77> 계속해서, 얻어진 혼성화 강도 정량 데이터로부터 유전자형 판별을 위한 입력 벡터 (input vector)를 계산한다(309). 유전자형 판별 함수 설정 방법에서 설명한 바와 마찬가지로 H-L 추정 방법을 적용하여 비율 성분을 계산한다. 유전자형 판별 함수 설정 시와

마찬가지로 MA그래프를 그릴 필요가 있을 경우에는 비율 성분과 함께 강도 성분도 계산한다.

<78> 마지막으로, 얻어진 입력 벡터를 판별 함수에 적용하여 유전자형을 판별한다(311). 타겟 핵산에 대한 테스트 결과와 표준 핵산에 대한 결과를 동일 MA 그래프 상에 도시는 것이 육안으로 확인할 수 있어서 바람직하다.

<79> 도 6은 MODY2 alpha 1 유전자의 엑손(exon) 4의 18번 돌연변이 위치(E04-18)에 대한 유전자형 판별 함수 설정을 위해 도기한 MA 그래프 위에 유전자형을 모르는 타겟 핵산의 테스트 결과를 함께 도기한 것이다. 도 4의 MA 그래프를 얻기 위한 실험 방법과 동일하게 진행하여 도 6의 MA 그래프를 도시하였다. 타겟 핵산의 혼성화 테스트 결과가 원(600)으로 표시한 영역에 도시되었다. 이 경우 타겟 핵산의 유전자를 정상형으로 판별할지 돌연변이형으로 판별할지를 결정해야 한다.

<80> 유전자형 판별은 다음과 같이 진행한다. 먼저, 추정된 회귀 계수(logistic regression coefficient)(a, b)를 가지는 판별 함수에 원(600)으로 표시된 부분의 m(비율 성분)을 입력 벡터로 적용한 후, 각 유전자형에 속할 사후 확률(posterior probability)을 계산한다.

<81> $P(\text{정상} | m(\text{비율 성분})) = \exp(a+bm) / \{1+\exp(a+bm)\}$

<82> $P(\text{돌연변이} | m(\text{비율 성분})) = 1 / \{1+\exp(a+bm)\} = 1 - P(\text{정상} | m(\text{비율 성분}))$

<83> 이어서, 각 유전자형에 속할 사후 확률 값 중에 큰 값을 가지는 값으로 유전자형을 판별한다. 바람직하기로는 각 유전자형에 속할 사후 확률 값 중에 큰 값을 가지는 값으로 잠정적으로 유전자형을 판별한 후, 상기 사후 확률 값의 특정 유의수준에 대한 신뢰

도를 확인하여 상기 신뢰도를 충족시키지 못할 경우 다시 유전자형에 대한 판별을 유보한다. 유전자형 판별 신뢰도 확인은 다음과 같이 진행한다. 먼저, 특정 유의수준에 대한 사후확률 값의 신뢰구간을 계산하고 신뢰구간이 0.5를 포함할 경우에는 유전자형 판별을 유보(nocall)한다. 즉, 유전자형을 그레이 존(gray zone)으로 설정한다. 사후확률의 신뢰구간 값을 계산하는 방법은 호스머 와 렘쇼(Hosmer, D.W, Jr. and Lemeshow, S)의 응용 로지스틱 회귀분석(Applied Logistic Regression, John Wiley & Sons, Inc, 1989) 1장에 자세히 기술되어 있다. 상기 저서의 내용은 참조용으로 본 명세서에 속하는 것으로 한다. 유전자형 판단을 보다 엄격하게 하는 방법으로는 신뢰구간이 0.5보다 큰 값, 예를 들어 소정 기준 값인 0.7을 포함할 때에도 유전자형 판별을 유보하는 것이다. 하지만 판별을 유보하는 경우가 너무 많으면 DNA 칩이 제대로 작동하지 않는 것이므로 소정 기준 값 설정은 유전자형 판별 유보율(nocall rate)과 유전자형 판별 오류율(miscall rate)을 통해 최적화하는 것이 필요하다.

<84> 한편, 강도 성분이 낮아질 때 비율 성분이 현저히 감소하는 경향을 보일 경우, 비율 성분이 감소하는 부분집합을 제외하고 나머지 비율 성분들의 집합을 학습세트로 사용한다. 이 학습세트에 대해서 로지스틱 회귀분석(logistic regression)을 판별함수로 사용하여 회귀 계수를 구한 경우에는 타겟 핵산의 혼성화 결과가 비율 성분이 감소하는 강도 성분 영역에 속할 경우에도 판별을 유보한다.

<85> 유전자형 판별 결과 보정

<86> 상술한 단계를 거쳐서 유전자형을 판별한 후, 판별 결과를 보정함으로써 판별의 유보나 판별 오류를 최소화할 수 있다. 교차-혼성화(cross-hybridization) 결과를 보고 유전자형 판별 결과를 보정할 수 있다. 예를 들어 특정한 위치에 돌연변이가 있는 표준 핵

산이 그 위치에 돌연변이가 있는지를 확인하기 위해 선발된 프로브 쌍이 아니라 다른 위치에 돌연변이가 있는지를 확인하기 위한 프로브 쌍과 혼성화되는 교차 혼성화 결과가 사전에 확인되면 이를 이용해서 최종 판별 결과를 보정한다.

<87> 구체적인 최종 판별 결과 보정의 예를 도 7 및 도 8을 참조하여 설명한다. 도 7 및 도 8의 실험 결과 또한 도 4의 실험 결과를 얻기 위한 방법과 동일하게 진행하여 얻었다. 도 7은 정상형으로 판별되어야 할 유전자가 돌연변이형으로 판별된 경우를, 도 8은 정상형으로 판별되어야 할 유전자의 유전자형 판별이 유보된 경우를 도시한 MA 그래프이다.

<88> 도 7은 MODY2 alpha 1 유전자의 엑손(exon) 7 의 16번 위치(E07-16)의 유전자형을 판별하기 위한 E07-16 프로브 쌍에 엑손 7과는 완전히 다른 엑손 8의 5번 위치(E08-05)에 변이가 있는 것으로 판별된 표준 핵산이 혼성화된 결과를 도시한 MA 그래프이다. 채워진 원(●)은 E07-16 정상형 표준 핵산과 프로브 쌍이 혼성화되었을 때의 결과이고, 채워진 세모(▲)는 E07-16 돌연변이형 표준 핵산과 프로브 쌍이 혼성화되었을 때의 결과이고, 비워진 원(○)으로 이루어진 타원(700)은 E08-05 위치 돌연변이 표준 핵산이 혼성화된 결과이다. 이 경우 E08-05 위치 돌연변이 표준 핵산은 E07-16 위치 돌연변이로 판별된다.

<89> 도 8은 MODY2 alpha 1 유전자의 엑손(exon) 8의 05번 위치(E08-05)에 돌연변이가 있는지를 확인하기 위한 E08-05 확인용 프로브 쌍에 E07-16 위치 돌연변이 표준 핵산과 E08-06 위치 돌연변이 표준 핵산이 혼성화된 결과를 동시에 도시한 MA 그래프이다. 채워진 원()은 E08-05 정상형 표준 핵산과 프로브 쌍이 혼성화되었

을 때의 결과이고, 채워진 세모(▲)는 E08-05 돌연변이형 표준 핵산과 프로브 쌍이 혼성화되었을 때의 결과이고, 비워진 세모(△)로 이루어진 타원(800)은 E08-06 위치 돌연변이 표준 핵산이 혼성화된 결과이고, 비워진 원(○)으로 이루어진 타원(900)은 E07-16 위치 돌연변이 표준 핵산이 혼성화된 결과이다. 이 경우, E07-16 위치 돌연변이 표준 핵산 뿐만 아니라 E08-06 위치 돌연변이 표준 핵산도 교차 혼성화가 일어나 유전자형 판별이 유보된다.

<90> 즉, 도 7에 도시되어 있는 바와 같이 E08-05 위치에 돌연변이가 있는 표준 핵산이 E08-05 위치는 물론 E07-16 위치에서도 돌연변이가 있는 것으로 판별되고, 도 8에 도시되어 있는 바와 같이 E07-16 위치에 돌연변이가 있는 표준 핵산이 교차 혼성화에 의해 E08-05 위치에서 유전자형 판별이 유보된다는 사실을 타겟 핵산의 혼성화 전에 알 수 있다면 다음과 같이 유전자형 판별 결과 보정이 가능하다.

<91> 즉, 타겟 핵산과 칩과의 혼성화 결과 타겟 핵산이 E08-05 위치 및 E07-16 위치에서 동시에 돌연변이가 있는 것으로 판별되면 E07-16 위치가 돌연변이 유전자로 판별되는 것은 교차 혼성화에 의한 것이므로, 이 타겟 핵산의 유전자형을 다시 E07-16 위치에서는 정상형으로 판별하도록 보정한다. 또한 타겟 핵산과 칩과의 혼성화 결과 E07-16 위치에서는 돌연변이로 E08-05 위치에서 유전자형 판별이 유보된다면 E08-05 위치에서는 정상형으로 판별하도록 보정한다.

<92> 한편, 도 8의 결과로부터 E07-16 위치 돌연변이 표준 핵산 뿐만 아니라 E08-06 위치 돌연변이 핵산도 E08-05 확인용 프로브 쌍과 결합했을 경우 교차 혼성화가 일어나 유전자형 판별을 유보하게 됨을 알았고, E08-06 확인용 프로브 쌍에서는 이와 같은 교차 혼성화가 생기지 않음을 확인하였다면 다음과 같이 유전자형 판별 보정이 가능하다.

<93> 즉, 타겟 핵산과 칩과의 혼성화 결과 E08-06 위치에서 돌연변이 유전자로 판별되고 E08-05 위치에서는 판별이 유보될 경우에는 이 타겟 핵산의 E08-05 위치 유전자형을 정상형 유전자로 판별하도록 보정한다.

【발명의 효과】

<94> 본 발명에 따르면, 강건하게 유전자형을 판별할 수 있다. 각 돌연변이 위치별로 최적의 프로브 쌍을 선별한 후 이를 이용하여 유전자형 판별을 시도하기 때문에 오류가 없게 유전자형을 판별할 수 있다. 또, 유전자형 판별 함수 설정을 위한 벡터를 통계학적으로 강건한 추정 방법인 H-L 추정을 사용하므로 실험에서의 오류에 영향을 덜 받을 수 있다. 또, 유전자형 판별의 오류가 치명적인 경우에는 유전자형 판별 과정에서 유전자형 판별 기준을 엄격하게 적용하여 판별의 정확성을 향상시킬 수 있다.

<95> 또, 본 발명에 따른 방법을 사용할 경우 QC 및 QA 확보가 가능하다. 동일 프로브를 반복해서 칩 상에 올려놓고 이들간에 변동을 살펴볼 수 있기 때문에 이상데이터(outlier data)에 대한 여과가 가능하고, 유전자형 판별 함수 설정 단계에서 정상형 표준 핵산과 정상형 유전자에 완전 일치하는 프로브들과의 혼성화를 반복해서 얻은 실험 결과를 바탕으로 칩 불량에 해당하는 기준을 설정할 수도 있다.

<96> 그리고, 하나의 칩 상에 놓여진 각 돌연변이 위치 당 프로브 쌍에 대한 교차 혼성화의 양상을 활용하여 유전자형 판별결과를 보정할 수 있다. 본 발명에서는 각 돌연변이 위치 당 하나의 프로브 쌍만 사용하기 때문에 교차-혼성화를 보다 명확하게 확인할 수 있다. 따라서 원래 정상형으로 판별되어야 할 유전자의 유전자형 판별이 유보되거나 돌

연변이형으로 판별될 경우 교차-혼성화의 결과를 사용하여 유전자형을 정상형으로 보정할 수 있다.

【특허청구범위】**【청구항 1】**

각 돌연변이 위치 별로 정상형 유전자에 완전히 일치하는 프로브와 돌연변이형 유전자에 완전히 일치하는 프로브로 이루어진 최적 프로브 쌍만이 배열되어 있는 DNA 칩을 사용하여 유전자형을 판별하는 유전자형 판별 방법.

【청구항 2】

제1 항에 있어서, 상기 각 돌연변이 위치 별로 상기 최적 프로브 쌍은 2회 이상 배열되어 있는 것을 특징으로 하는 유전자형 판별 방법.

【청구항 3】

제2 항에 있어서, 상기 각 돌연변이 위치 별로 상기 정상형 유전자에 완전히 일치하는 프로브가 2회 이상 배열되고 그 옆에 상기 돌연변이형 유전자에 완전히 일치하는 프로브가 2회 이상 배열되어 있는 것을 특징으로 하는 유전자형 판별 방법.

【청구항 4】

제1 항에 있어서, 상기 최적의 프로브 쌍은

인 실리코 방법으로 정상형 유전자에 완전히 일치하는 프로브와 돌연변이형 유전자에 완전히 일치하는 프로브로 이루어진 복수개의 서로 다른 프로브 쌍을 설계하는 단계;

기관 상에 상기 복수개의 서로 다른 프로브 쌍을 고정화하여 최적 프로브 쌍 선발용 칩을 형성하는 단계;

상기 최적 프로브 쌍 선발용 칩에 표준 핵산을 혼성화시키는 단계;

혼성화 강도 정량 데이터를 수집하는 단계; 및

상기 혼성화 강도 정량 데이터를 사용하여 하기 식의 계산 값이 가장 큰 프로브 쌍을 선발하는 단계를 포함하는 최적 프로브 쌍 선발 방법에 의해 선발된 것을 특징으로 하는 유전자형 판별 방법.

$$\{\text{Mean}(\ln(r^{\text{wt}})) - 2 \text{SD}(\ln(r^{\text{wt}}))/\sqrt{N^{\text{wt}}}\} - \{\text{Mean}(\ln(r^{\text{mt}})) + 2\text{SD}(\ln(r^{\text{mt}}))/\sqrt{N^{\text{mt}}}\}$$

상기 식에서, N은 각 표준 핵산과 혼성화를 시행한 횟수이고, 상기 r^{wt} 은 정상형 표준 핵산의 상기 정상형 유전자에 완전히 일치하는 프로브에 대한 혼성화 강도를 상기 정상형 표준 핵산의 상기 돌연변이형 유전자에 완전히 일치하는 프로브에 대한 혼성화 강도로 나눈 값이고, 상기 r^{mt} 는 돌연변이형 표준 핵산의 상기 정상형 유전자에 완전히 일치하는 프로브에 대한 혼성화 강도를 상기 돌연변이형 표준 핵산의 상기 돌연변이형 유전자에 완전히 일치하는 프로브에 대한 혼성화 강도로 나눈 값이고, Mean은 평균값을, SD는 표준편차를 나타낸다.

【청구항 5】

제1 항에 있어서, 상기 유전자형 판별 방법은

(a) 상기 DNA 칩에 유전자형을 이미 알고 있는 표준 핵산을 혼성화시켜 얻어진 데이터로부터 유전자형 판별 함수를 설정하는 단계; 및

(b) 상기 DNA 칩에 유전자형을 모르는 타겟 핵산을 혼성화시켜 얻어진 데이터로부터 입력 벡터를 계산하고 상기 입력 벡터를 상기 유전자형 판별 함수에 입력하여 상기 타겟 핵산의 유전자형을 판별하는 단계를 포함하는 것을 특징으로 하는 유전자형 판별 방법.

【청구항 6】

제5 항에 있어서, 상기 (a) 단계는,

(a-1) 상기 DAN 칩에 유전자형을 이미 알고 있는 표준 핵산을 혼성화시켜 얻어진 혼성화 강도 정량 데이터를 수집하는 단계;

(a-2) 상기 정상형 유전자에 완전히 일치하는 프로브의 혼성화 강도와 상기 돌연변이형 유전자에 완전히 일치하는 프로브의 혼성화 강도의 모든 가능한 조합의 비율 값을 계산한 후, 호지-레만 추정에 의해 이 비율 값들 중 중위수를 구하여 이 중위수에 로그를 취해서 판별 함수 설정용 벡터의 비율 성분을 계산하는 단계; 및

(a-3) 상기 (a-1) 및 (a-2) 단계를 복수개의 상기 DNA 칩에 대하여 반복하여 얻어진 상기 벡터들의 집합을 사용하여 상기 유전자형 판별 함수를 설정하는 단계를 포함하는 것을 특징으로 하는 유전자형 판별 방법.

【청구항 7】

제 6항에 있어서, 상기 (a-3) 단계는

상기 벡터들의 집합을 사용하여 상기 판별 함수인 로지스틱 회귀분석의 회귀 계수를 구하는 단계인 것을 특징으로 하는 유전자형 판별 방법.

【청구항 8】

제6 항 또는 제 7항에 있어서, 상기 (a-2) 단계는 상기 비율 성분과 함께 상기 정상형 유전자에 완전히 일치하는 프로브의 혼성화 강도와 상기 돌연변이형 유전자에 완전히 일치하는 프로브의 혼성화 강도의 모든 가능한 조합의 곱의 값을 계산한 후, 호지-레만 추정에 의해 이 곱의 값들 중 중위수를 구하여 이 중위수에 로그를 취해서 2로 나누어서 상기 판별 함수 설정용 벡터의 강도 성분도 함께 계산하는 단계이고,

상기 (a-3) 단계 전에, 상기 비율 성분을 y축에 상기 강도 성분을 x축에 도시하는 단계를 더 구비하고,

상기 (a-3) 단계는 상기 도시된 그래프에서 상기 강도 성분이 변화하더라도 상기 비율 성분이 변화하지 않을 경우에는 상기 비율 성분의 전체를, 상기 강도 성분이 변화함에 따라 상기 비율 성분이 변하는 경우에는 상기 비율 성분이 변화하는 부분을 제외하고 나머지 비율 성분들만을 사용하여 상기 판별 함수를 설정하는 단계인 것을 특징으로 하는 유전자형 판별 방법.

【청구항 9】

제6 항 또는 제 7항에 있어서, 상기 (a-2) 단계는 상기 비율 성분과 함께 상기 정상형 유전자에 완전히 일치하는 프로브의 혼성화 강도와 상기 돌연변이형 유전자에 완전히 일치하는 프로브의 혼성화 강도 중의 최대값을 모든 가능한 조합에 대해 계산한 후, 호지-레만 추정에 의해 이 최대 값들 중 중위수를 구하여 이 중위수에 로그를 취해서 판별 함수 설정용 벡터의 강도 성분도 함께 계산하고,

상기 (a-3) 단계 전에, 상기 비율 성분을 y축에 상기 강도 성분을 x축에 도시하는 단계를 더 구비하고,

상기 (a-3) 단계는 상기 도시된 그래프에서 상기 강도 성분이 변화하더라도 상기 비율 성분이 변화하지 않을 경우에는 상기 비율 성분의 전체를, 상기 강도 성분이 변화함에 따라 상기 비율 성분이 변하는 경우에는 상기 비율 성분이 변화하는 부분을 제외하고 나머지 비율 성분들만을 사용하여 상기 판별 함수를 설정하는 단계인 것을 특징으로 하는 유전자형 판별 방법.

【청구항 10】

제 6항에 있어서, 상기 (a-2) 단계 전에

상기 혼성화 강도 정량화 데이터 중 스왑 지름 유효값의 임계치보다 큰 지름을 가지는 불량 스왑의 혼성화 강도 정량화 데이터를 제거하는 단계를 더 포함하는 것을 특징으로 하는 유전자형 판별 방법.

【청구항 11】

제 5항에 있어서, 상기 (b) 단계는

(b-1) 상기 DNA 칩에 유전자형을 모르는 타겟 핵산을 혼성화시켜 얻어진 혼성화 강도 정량화 데이터를 수집하는 단계;

(b-2) 상기 정상형 유전자에 완전히 일치하는 프로브의 혼성화 강도와 돌연변이형 유전자에 완전히 일치하는 프로브의 혼성화 강도의 모든 가능한 조합의 비율 값을 계산한 후, 호지-레만 추정에 의해 이 비율 값들 중 중위수를 구하여 이 중위수에 로그를 취해서 유전자형 판별을 위한 입력 벡터를 계산하는 단계; 및



(b-3) 상기 입력 벡터를 상기 판별 함수에 입력하여 유전자형을 판별하는 단계를 포함하는 것을 특징으로 하는 유전자형 판별 방법.

【청구항 12】

제 11항에 있어서, 상기 (b-3) 단계는

상기 입력 벡터를 상기 판별 함수에 입력한 후 각 유전자형에 속할 사후 확률 값을 계산하여 상기 사후 확률 값이 큰 유전자형으로 판별하는 단계를 포함하는 것을 특징으로 하는 유전자형 판별 방법.

【청구항 13】

제 11항에 있어서, 상기 (b-3) 단계는

상기 입력 벡터를 상기 판별 함수에 입력한 후 각 유전자형에 속할 사후 확률 값을 계산하여 상기 사후 확률 값이 큰 유전자형으로 잠정적으로 판별하는 단계; 및

상기 사후 확률 값의 특정 유의수준에 대한 신뢰도를 확인하여 상기 신뢰도를 충족시키지 못할 경우 다시 유전자형에 대한 판별을 유보하는 단계를 포함하는 것을 특징으로 하는 유전자형 판별 방법.

【청구항 14】

제 11항에 있어서, 상기 (b-3) 단계 전에

상기 혼성화 강도 정량화 데이터 중 스왑 지름 유효값의 임계치보다 큰 지름을 가지는 불량 스왑의 혼성화 강도 정량화 데이터를 제거하는 단계를 더 포함하는 것을 특징으로 하는 유전자형 판별 방법.



【청구항 15】

제 5항에 있어서, 상기 (b) 단계 후에,

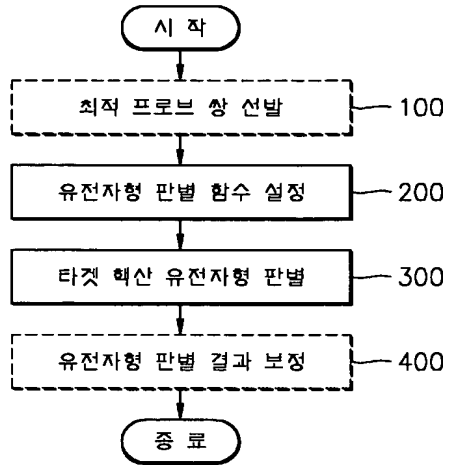
상기 각 돌연변이 위치 별로 교차 혼성화 결과를 체크하여 유전자형 판별 결과를 보정하는 단계를 더 포함하는 것을 특징으로 하는 유전자형 판별 방법.

【청구항 16】

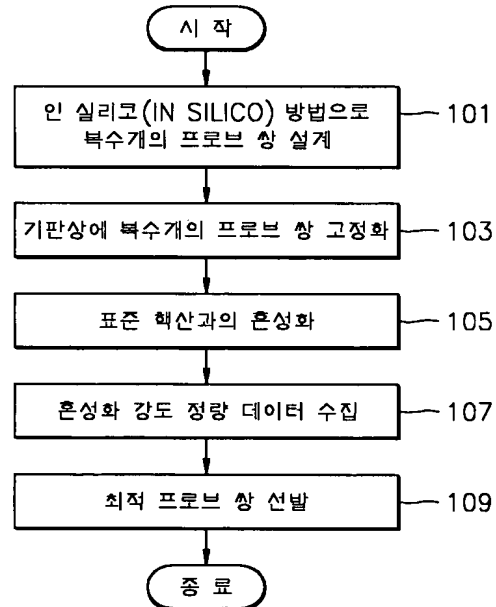
제1 항 내지 제7 항 및 제 10항 내지 제 15항 중 어느 한 항에 있어서, 상기 유전자형 판별 방법에 사용되는 DNA 칩.

【도면】

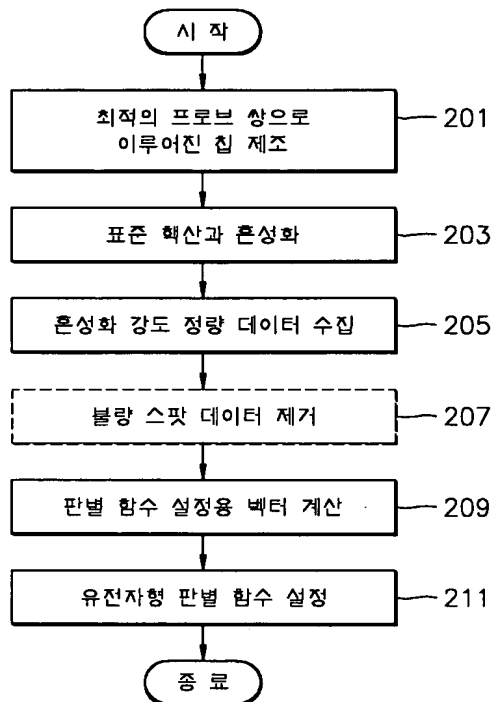
【도 1】



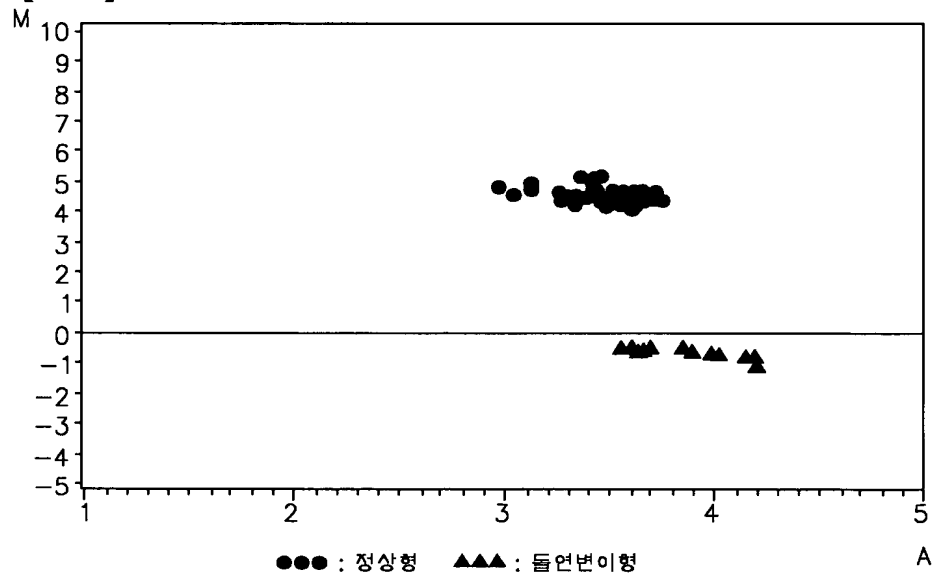
【도 2】



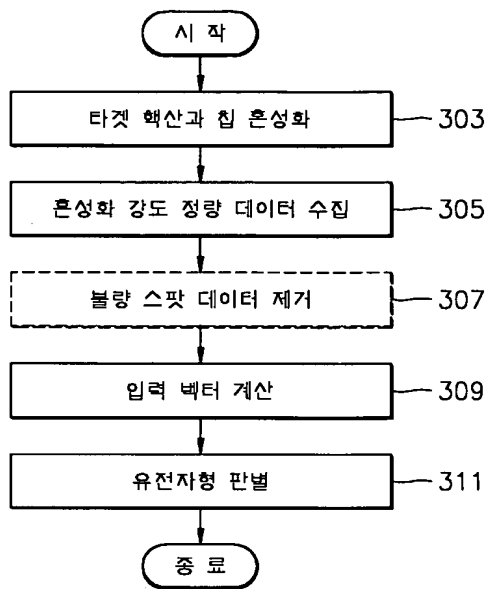
【도 3】



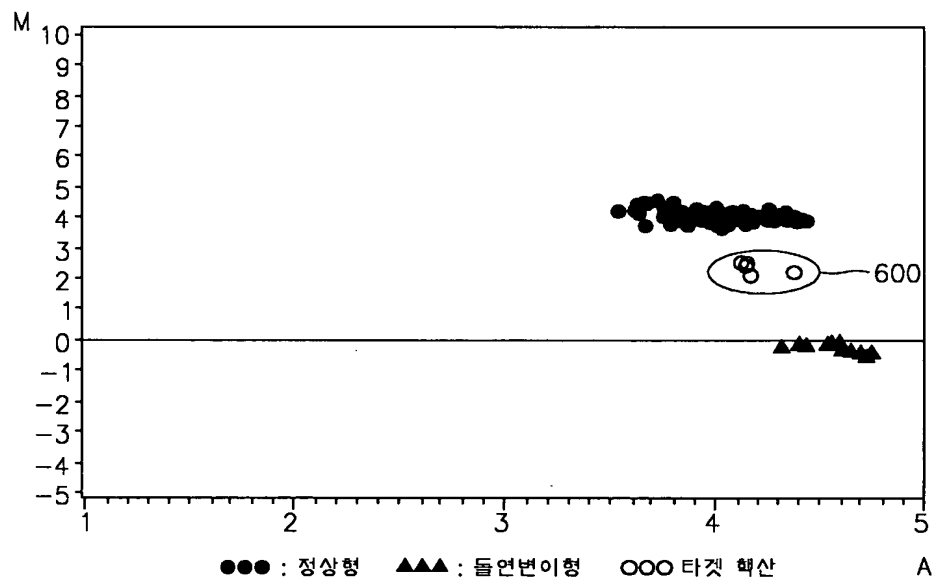
【도 4】



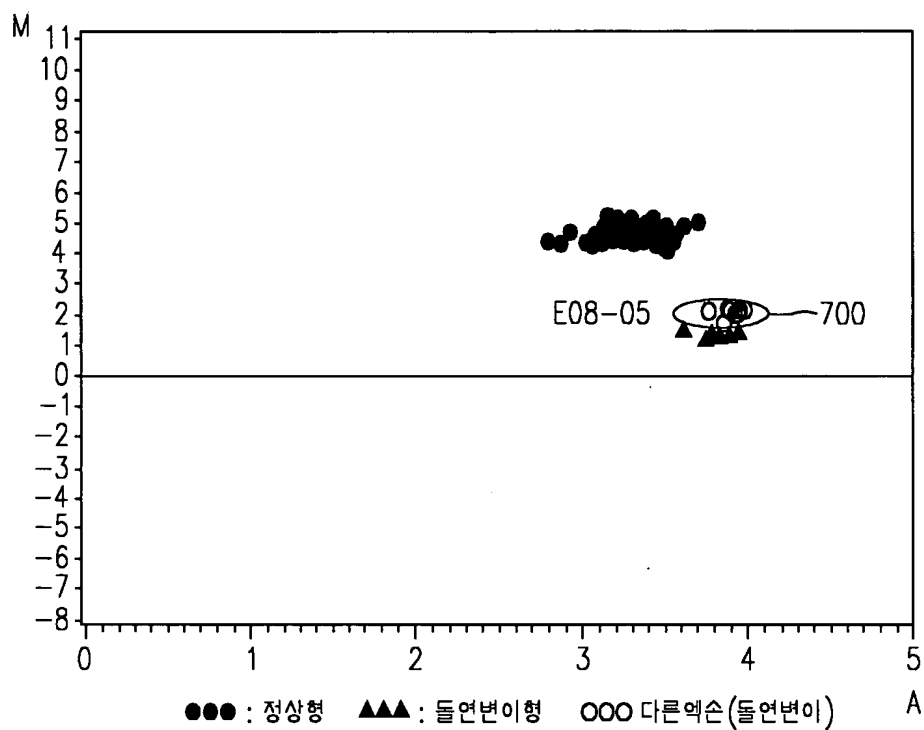
【도 5】



【도 6】



【도 7】



【도 8】

